Claims

What is claimed is:

1. A system for extracting information from network data, comprising:

an input interface connected to at least one source of network data; and

a network event sensor, communicating with the input interface, the network event sensor applying at least a lexical engine to the network data to identify at least one network event.

2. The system of claim 1, wherein the at least one source of network data comprises an observation port connected to a network and continuously capturing network data from the network.

3. The system of claim 2, wherein the observation port comprises a network interface card.

4. The system of claim 3, wherein the network comprises at least one of an Ethernet network, a token ring network, and a TCP/IP network.

5. The system of claim 3, wherein the network interface card is invisible to the network.

6. The system of claim 1, wherein the at least one source of network data comprises stored network data.

7. The system of claim 6, wherein the stored network data comprise at least one of captured network files, Website mirrors, archives of Usenet files, and archives of email files.

8. The system of claim 1, further comprising an interpreter module, the interpreter module scanning the network data to generate logical groupings of the network data.

9. The system of claim 8, wherein the logical groupings comprise packets.

10. The system of claim 8, wherein the interpreter module removes low-level encoding information from the network data to generate the logical groupings.

11. The system of claim 10, wherein the low-level encoding information removed by the interpreter module comprises hardware addressing information.

12. The system of claim 8, further comprising an assembler module, communicating with the interpreter module, the assembler module scanning the logical groupings to generate at least one session object.

13. The system of claim 12, wherein the at least one session object comprises at least one session file.

14. The system of claim 12, wherein the assembler module scans the logical groupings by examining at least one of source address, destination address, sequence numbers, source port, and destination port to generate the at least one session object.

15. The system of claim 12, wherein the network event sensor applies the lexical engine to the at least one session object to identify the at least one network event as at least one of a predetermined set of event types.

16. The system of claim 15, wherein the lexical engine detects the presence of at least one predefined keyword to identify the at least one of a predetermined set of event types.

17. The system of claim 16, wherein the predetermined set of event types comprises at least one of TCP, IP, UDP, SMTP, HTTP, NNTP, FTP, TELNET, DNS, RIP, BGP, MAIL, NEWS, HTML, XML, PGP, S/MIME, POP, IMAP, V-CARD, ICMP, NetBUI, IPX and SPX.

18. The system of claim 16, wherein the lexical engine accumulates a total number of occurrences for the at least one predefined keyword to identify the event type.

19. The system of claim 18, wherein the lexical engine applies a threshold to the number of occurrences to identify the event type.

20. The system of claim 12, wherein the network event sensor applies the lexical engine recursively to identify more than one event type contained in the at least one session object.

21. The system of claim 15, further comprising an extractor module, the extractor module extracting the at least one network event from the at least one session object according to the at least one of a predetermined set of event types.

22. The system of claim 21, wherein the extractor module comprises a library of extractor types, each of the extractor types corresponding to at least one of the at least one of a predetermined set of event types.

23. The system of claim 22, wherein the extractor module stores a minimum subset of the network data to reconstruct the at least one network event.

24. The system of claim 23, wherein the minimum subset of the network data is stored in a database.

25. The system of claim 24, further comprising a presentation module, communicating with the database, the presentation module querying the database for information related to the at least one network event.

26. The system of claim 1, wherein the network event sensor also applies a port detection engine to the network data to identify the at least one network event.

27. The system of claim 1, wherein the at least one source of network data comprises a plurality of sources of network data.

28. A method for extracting information from network data, comprising the steps of:

a) receiving network data from at least one source of network data; and

b) applying at least a lexical engine to the network data to identify at least one network event.

29. The method of claim 28, wherein the at least one source of network data comprises an observation port connected to a network and continuously capturing network data from the network.

30. The method of claim 29, wherein the observation port comprises a network interface card.

31. The method of claim 30, wherein the network comprises at least one of an Ethernet network, a token ring network, and a TCP/IP network.

32. The method of claim 30, wherein the network interface card is invisible to the network.

33. The method of claim 28, wherein the at least one source of network data comprises stored network data.

34. The method of claim 33, wherein the stored network data comprise at least one of captured network files, Website mirrors, archives of Usenet files, and archives of email files.

35. The method of claim 28, further comprising a step of c) scanning the network data to generate logical groupings of the network data.

36. The method of claim 35, wherein the logical groupings comprise packets.

37. The method of claim 35, further comprising a step of d) removing low-level encoding information from the network data to generate the logical groupings.

38. The method of claim 37, wherein the low-level encoding information comprises hardware addressing information.

39. The method of claim 35, further comprising a step of e) scanning the logical groupings to generate at least one session object.

40. The method of claim 39, wherein the at least one session object comprises at least one session file.

41. The method of claim 39, wherein the step (e) of scanning the logical groupings comprises a step of f) examining at least one of source address, destination address, sequence numbers, source port, and destination port to generate the at least one session object.

42. The method of claim 39, further comprising a step of g) identifying the at least one network event as at least one of a predetermined set of event types.

43. The method of claim 42, wherein the step (g) of identifying comprises a step of (h) detecting the presence of at least one predefined keyword to identify the at least one of a predetermined set of event types.

44. The method of claim 43, wherein the predetermined set of event types comprises at least one of TCP, IP, UDP, SMTP, HTTP, NNTP, FTP, TELNET, DNS, RIP, BGP, MAIL, NEWS, HTML, XML, PGP, S/MIME, POP, IMAP, V-CARD, ICMP, NetBUI, IPX and SPX.

45. The method of claim 43, wherein the step (h) of detecting comprises a step of (i) accumulating a total number of occurrences for the at least one predefined keyword to identify the event type.

46. The method of claim 45, wherein the step (h) of detecting comprises a step (j) of applying a threshold to the number of occurrences to identify the event type.

47. The method of claim 39, wherein the step of b) applying at least the lexical engine comprises a step of k) applying the lexical engine recursively to identify more than one event type contained in the at least one session object.

48. The method of claim 42, further comprising a step of l) extracting the at least one network event from the at least one session object according to the at least one of a predetermined set of event types.

49. The method of claim 48, wherein the step (l) of extracting comprises a step of m) selecting at least one extractor module from a library of extractor types, each of the extractor types corresponding to at least one of the at least one of a predetermined set of event types.

50. The method of claim 49, further comprising a step of n) storing a minimum subset of the network data to reconstruct the at least one network event.

51. The method of claim 50, wherein the step (n) of storing comprises a step o) of storing the minimum subset of the network data in a database.

52. The method of claim 51, further comprising a step of p) querying the database for information related to the at least one network event.

53. The method of claim 28, further comprising a step q) of applying a port detection engine to the network data to identify the at least one network event.

54. The method of claim 28, wherein the at least one source of network data comprises a plurality of sources of network data.